



Saramago, P., Claxton, K., Welton, N. J., & Soares, M. (2020). Bayesian econometric modelling of observational data for cost-effectiveness analysis: Establishing the value of Negative Pressure Wound Therapy in the healing of open surgical wounds. *Journal of the Royal Statistical Society: Series A*, 183(4), 1575-1593.  
<https://doi.org/10.1111/rssa.12596>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1111/rssa.12596](https://doi.org/10.1111/rssa.12596)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley at <https://doi.org/10.1111/rssa.12596>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Bayesian econometric modelling of observational data for cost-effectiveness analysis: establishing the value of negative pressure wound therapy in the healing of open surgical wounds

Pedro Saramago and Karl Claxton,  
*University of York, UK*

Nicky J. Welton  
*University of Bristol, UK*

and Marta Soares  
*University of York, UK*

[Received September 2018. Revised June 2020]

**Summary.** In the absence of evidence from randomized controlled trials on the relative effectiveness of treatments, cost-effectiveness analyses increasingly use observational data instead. Treatment assignment is not, however, randomized, and naive estimates of the treatment effect may be biased. To deal with this bias, one may need to adjust for observed and unobserved confounders. In this work we explore and discuss the challenges of these adjustment strategies within a case-study of negative pressure wound therapy (NPWT) for the treatment of surgical wounds healing by secondary intention. We could not demonstrate that existing uncontrolled confounding affects NPWT effectiveness, and thus there was no evidence that NPWT was cost effective compared with standard dressings for the treatment of surgical wounds healing by secondary intention.

**Keywords:** Economic evaluation; Instrumental variables; Observational data; Statistical methods; Surgical wounds

## 1. Introduction

System level decisions regarding the use of healthcare technologies are often informed by clinical and cost-effectiveness evidence, relying crucially on evidence from randomized controlled trials that the technology improves health outcomes. Observational studies are, however, increasingly recommended to support such decisions where, after looking at randomized controlled trial evidence, significant clinical uncertainty remains—for example, the Cancer Drugs Fund in the UK (NHS England, 2019) or the ‘Surveillance, epidemiology, and end results program’ in the USA (National Cancer Institute, 2019)—or to provide crucial intelligence on technologies such as medical devices that face no regulatory requirement to demonstrate efficacy (European Medicines Agency, 2020) and thus often present with limited effectiveness evidence (Akhmetov and Bubnov, 2015; Henshall *et al.*, 2011; Byron *et al.*, 2014).

*Address for correspondence:* Pedro Saramago, Centre for Health Economics, University of York, Alcuin College, York, YO10 5DD, UK.  
E-mail: pedro.saramago@york.ac.uk

© 2020 The Authors. Journal of the Royal Statistical Society: Series A (Statistics in Society) 0964–1998/20/1831575 published by John Wiley & Sons Ltd on behalf of Royal Statistical Society. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

But non-randomized observational evidence means that there is no control over assignment. In this situation treatment groups may differ in aspects that are related to the outcome (prognostic variables) generating potential for confounding. There is little guidance on the analysis of observational data: it is presented either as a description of methods (Faria *et al.*, 2015) or generic checklists or questionnaires (Berger *et al.*, 2014; Kreif *et al.*, 2013; Motheral *et al.*, 2003). Where there is a good understanding of the process of selection into treatment, and this information has been collected, analyses may attempt to adjust for imbalances, e.g. using regression adjustment, propensity scores or matching. However, these methods cannot rule out the possibility of confounding on prognostic factors that have not been observed, and hence causality cannot be established. Alternative adjustment methods use instrumental variables (IVs) (Cameron and Trivedi, 2005; Greene, 2012; Wooldridge, 2013), variables that are associated with treatment assignment but not directly with health outcomes (only through treatment assignment). The use of instruments enables the identification of the effects of 'pure' changes in treatment assignment (not related to confounders), and in this way to estimate causal effects. Only a few applications have been published to date on the use of IV methods in health technology assessment (Hadley *et al.*, 2003; Brookhart *et al.*, 2006a, b; Schneeweiss *et al.*, 2008; Prentice *et al.*, 2014).

This paper explores the use of an existing observational cohort study (Chetter *et al.*, 2019) to establish the clinical and cost effectiveness of a medical device that is widely used in the National Health Service for surgical wound healing by secondary intention (SWHSI) called negative pressure wound therapy (NPWT).

Section 2 gives a description of the case-study and a summary of the observational data, together with the methods of analyses. The (causal) cost-effectiveness model is described, to provide the framework by which potential improvements in time to wound healing by NPWT impact on health-related quality of life (HRQOL) and costs. We also describe the specific regression methods that are used to determine the effectiveness of NPWT on time to wound healing, using adjustments for observed factors and IV estimation. Section 3 gives a description of the results of analyses for effectiveness, costs and HRQOL. The paper finishes with a discussion of the challenges that are encountered (Section 4) that can be generalized to the use of IVs in other health technology assessment contexts.

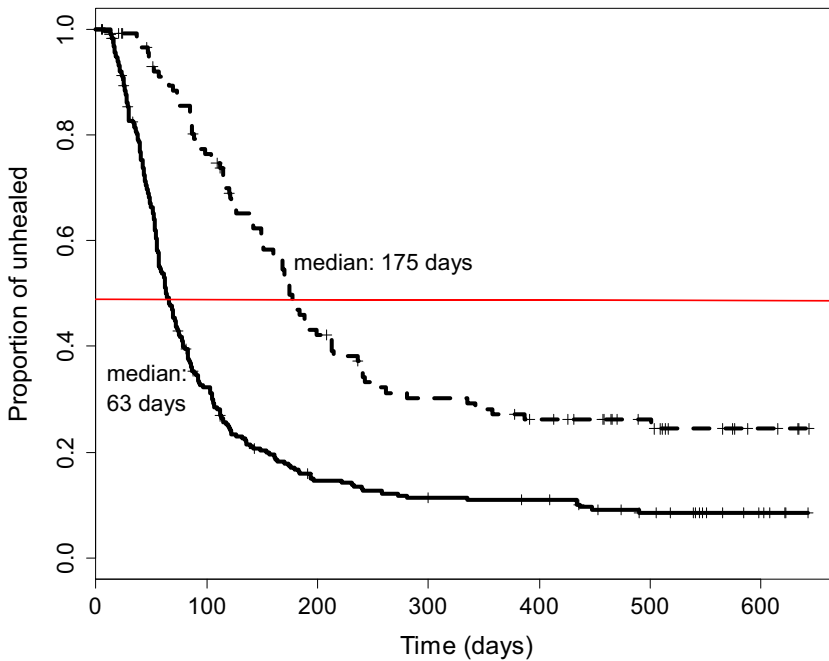
## 2. Methods

### 2.1. Description of the case-study and observational data available

NPWT is a medical device that is applied to the wound surface and creates a suction force (or vacuum) removing tissue fluid away from the treated wound area into a canister. NPWT has been claimed (Banwell and Musgrave, 2004) to speed up healing (the outcome that is most valued by patients (Cullum *et al.*, 2015)). However, there is no conclusive randomized control trial evidence of such an effect (Berger *et al.*, 2014) and, in light of this, the National Institute for Health and Care Excellence in 2008 did not reject NPWT but recommended further research (Leaper *et al.*, 2008; National Institute for Health and Care Excellence, 2008). Since, however, there has been no new relevant evidence (National Institute for Health and Care Excellence, 2011a, b) and, in all this time, NPWT has been widely used by health systems with no insight into its value for money.

#### 2.1.1. Summary description of the data

The largest and most comprehensive study to date in this population is a recent observational



**Fig. 1.** Time to healing: descriptive Kaplan–Meier curve: — — —, NPWT patients; ———, other treatments (non-NPWT patients)

cohort study that was undertaken in the UK (Chetter *et al.*, 2019). This study was longitudinal, conducted over a 21-month period (from February 18th, 2013, to November 30th, 2014) in eight study sites across Yorkshire, UK. Participants were recruited with incident surgical wounds requiring treatment (the inclusion and exclusion criteria have been reported elsewhere (Chetter *et al.*, 2019)) and were aimed to be followed up for 12 months or more.

The study recruited 393 participants and followed them up for 499 days, on average. Out of the total 393 participants, 22 died (5.6%) and 320 (81%) healed within the study. The mean time to healing of those who healed was 99 days. A full list of summary characteristics can be found in the on-line supplementary material (appendix 1, Table A1.1). 115 participants (29%) were treated with NPWT at some point during follow-up. Of those receiving NPWT only 69% (79 participants) healed within the study compared with 87% healed in those people who never received NPWT. Patients who received NPWT took longer to heal (a median time to healing of 175 days for NPWT *versus* 63 days without)—as highlighted in the Kaplan–Meier curve in Fig. 1. Given the observational nature of the study, imbalances in important prognostic variables (observed or unobserved) can justify the worse outcomes. The following subsection examines observed imbalances.

### 2.1.2. Imbalances and overlap between the groups

The majority of NPWT patients had a large (bigger than 25 cm<sup>2</sup>) wound compared with only 12% of non-NPWT patients and were more likely to have experienced skin and subcutaneous tissue loss whereas patients who did not receive NPWT had wounds involving skin only. These and other identified differences demonstrate that there is the potential for selection bias in the cohort data.

For a set of baseline covariates, Fig. 2(a) shows the mean differences between groups (NPWT minus non-NPWT), standardized (by dividing the differences by the standard deviation of the whole sample). Although there may be differences between the groups across a multitude of factors, none of the covariates analysed shows a difference bigger than 1 standard deviation across groups. The distribution of values for each covariate was also compared across groups to assess whether there were regions of non-overlap (an example plot is shown in Fig. 2(b) for the wound area). We verified that there was always some overlap, which indicates that the mechanism of selection of patients into the treatment groups is not clearly determined.

### 2.1.3. *Participants for whom an event (healing) was not observed (hard to heal)*

There were 73 patients who did not heal within the study, introducing censoring in the data. These patients were followed up on average for 416 days which, compared with average time to healing in those that healed of 99 days, suggests that these are ‘hard-to-heal’ patients. Also, these patients differed in their baseline characteristics from the patients who healed; for example, there were 23% more diabetics and 24% more patients with larger wounds, among others. A full list of summary characteristics by healing status can be found in appendix 1, Table A1.2, in the on-line supplementary material. This implies that censoring cannot be assumed uninformative which needs to be acknowledged in the methods of analyses (see Section 2.4).

## 2.2. *Causal model for determining cost effectiveness*

The outcome of analyses in determining cost effectiveness was the QALY: quality-adjusted life years gained (National Institute for Health and Care Excellence, 2013). QALYs weight lifetime lived for the HRQOL that the years are lived in. A 1-year time horizon was used and the UK National Health Service perspective was adopted for costs. To establish cost effectiveness it is important that the costs and HRQOL are based on the adjusted causal effect of NPWT on time to healing,  $\Delta t$ , and not the observed, unadjusted, effect. The cost-effectiveness model (equations (1)–(4)) computes the cost and HRQOL implications of treatment with NPWT. Equation (1) details that differences in QALY,  $\Delta Q$ , are evaluated from  $\Delta t$  and from  $u_h$  and  $u_{unh}$ , the HRQOL weights of healed and unhealed patients respectively:

$$\Delta Q = \Delta t(u_h - u_{unh}), \quad (1)$$

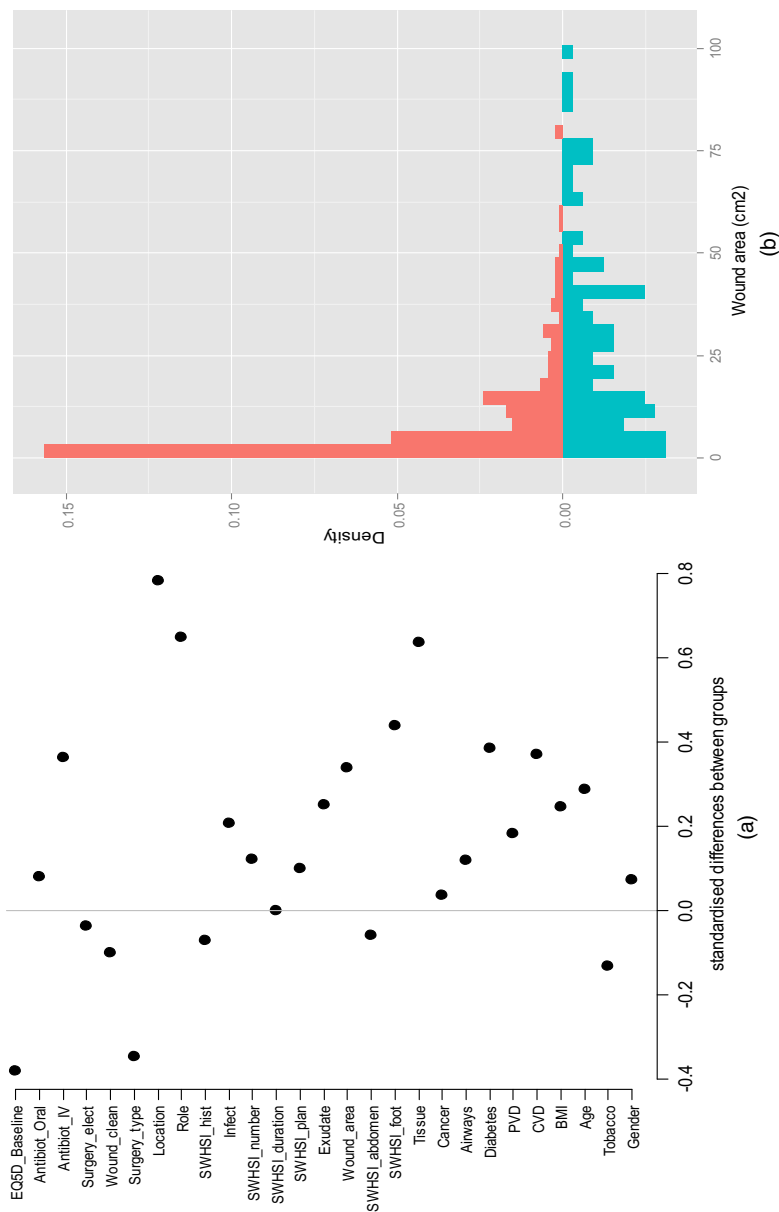
$$\Delta C = \Delta C_{dis} + \Delta C_t, \quad (2)$$

$$\Delta C_{dis} = \Delta t(c_{unh} - c_h), \quad (3)$$

$$\Delta C_t = \underbrace{c_1 \text{ ToT} + c_0(\Delta t - \text{ToT})}_{\text{NPWT group}} - \underbrace{c_0 \Delta t}_{\text{non-NPWT group}}. \quad (4)$$

The difference in costs,  $\Delta C$ , is broken down into disease-related costs,  $\Delta C_{dis}$ , and the costs of the treatments themselves,  $\Delta C_t$  (equation (2)).  $\Delta C_{dis}$  are determined by applying  $\Delta t$  to the difference in the costs of unhealed and healed patients,  $c_{unh}$  and  $c_h$  respectively (equation (3)).  $\Delta C_t$  (equation (4)) assumes that patients receive some form of treatment up to healing. The treatment costs of the ‘non-NPWT group’ is simply the average cost of dressings,  $c_0$ , multiplied by  $\Delta t$ . For patients receiving NPWT, the average daily cost of NPWT,  $c_1$ , is applied to the mean time on NPWT treatment, TOT, and in the remaining time that patients are assumed to use dressings (with a daily cost equal to  $c_0$ ; see equation (4)).

The description of the model above identifies evidence requirements, and the subsections ahead detail the estimation of the causal effect of NPWT on time to healing and inferences on



**Fig. 2.** (a) Standardized mean differences between treatment groups and (b) histogram of wound area for each treatment group (■, patients on treatments other than NPWT; ■, patients on NPWT treatment); EQ5D\_Baseline, EQ-5D index value at baseline; Antibiot\_Oral, antibiotics treatment oral; Antibiot\_IV, antibiotics treatment intravenous; Surgery\_elect, emergency surgery (versus elective); Wound\_clean, wound cleanliness, dirty (versus clean and clean or contaminated); Surgery\_type, surgery type, colorectal (versus others); Location, treatment location, inpatient (versus outpatient); Role, role of health professional, doctor (versus nurse); SWHSI\_hist, history of SWHSI; Infect, infection wound at baseline; SWHSI\_number, number of SWHSI; SWHSI\_duration, SWHSI duration at baseline (in days); SWHSI\_plan, wound type, dehiscent (versus planned); Exudate, Exudate level, low (versus medium or high); Wound\_area, wound area (cm²); SWHSI\_abdomen, SWHSI location, abdomen (versus other locations); SWHSI\_foot, SWHSI location, foot (versus other locations); Tissue, tissue involvement, skin and subcutaneous tissue loss (versus skin loss); Cancer, patient with cancer; Airways, patient with chronic obstructive airways or pulmonary disease; Diabetes, patient with diabetes; PVD, patients with peripheral vascular disease; CVD, cardiovascular disease; BMI, body mass index; Age, age in years; Tobacco, tobacco use, smoker; Gender, male (versus female); npwt, patients on NPWT; no npwt, patients on treatments other than NPWT

the effects of healing on HRQOL weights and costs. The remaining parameters (ToT,  $c_0$  and  $c_1$ ) were informed by the means observed in the data set.

### 2.3. Implementation and software

A Bayesian approach was used throughout to allow for the uncertainty in predicting treatment allocation to be fully considered and thus correctly reflected in decision uncertainty. Analyses were implemented in R (R Development Core Team, 2013) and inferences in WinBugs (Lunn *et al.*, 2000). Model selection was based on the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002; Gelman, 2014; Lunn, 2013), which reflects a trade-off between model fit and model complexity. DIC differences of 5 were considered meaningful (Spiegelhalter *et al.*, 2002, 2014). Econometric tests of the relevance and validity of instruments are not developed in the Bayesian context, and frequentist methods were thus implemented in Stata (StataCorp, 2013). Decision uncertainty was reported by using the probability that each intervention was cost effective.

### 2.4. Inferences on time to healing

In cost effectiveness, where expected values are of interest, time-to-event outcomes are traditionally analysed by using parametric survival analysis. In this case-study, standard parametric analyses would struggle to provide a good fit because of the asymptote that is observed in Fig. 1, and more complex, and therefore flexible, functions (such as splines) would need to be used. However, these are not readily available in software packages that are used for Bayesian analyses (such as WinBugs (Lunn, 2013; Lunn *et al.*, 2000)). For this reason, we took a different approach. We first generated a complete data set, enabling the use of linear regression in which IV methods are well established. We acknowledge that linear regression assumes that times to event (conditional on the explanatory variables) are normally distributed, which is unlikely because of the right skewness of these data. We therefore extend the analyses to consider more explicitly the hard-to-heal patients, who account for some degree of skewness.

#### 2.4.1. Using a complete data set

To generate a complete data set, censored times to healing were imputed by assuming that patients healed the day after they were censored. This will closely reflect the data but, as it is unlikely that all individuals healed so close to censoring, it will systematically underestimate healing times. Time to healing (the index  $i$  for patients is omitted from the text throughout),  $y$ , was then regressed on treatment,  $x$ , and a set of relevant adjustment factors that have been observed,  $\mathbf{A}$  (the design matrix). The model that was implemented is described in equation (5):

$$\begin{aligned} y_i &\sim N(y_i^*, \sigma), \\ y_i^* &= \mu_1 + \mu_2 \cdot x_i + \boldsymbol{\mu}^{\text{adj}} \cdot \mathbf{A}_i \quad \text{for } i = 1, \dots, n. \end{aligned} \quad (5)$$

When IVs are considered, a two-stage regression is needed (equation (6)). The first stage predicts treatment allocation conditional on the instrument(s),  $\mathbf{z}$ , and a set of relevant covariates,  $\mathbf{B}$ . The predicted probabilities  $p^*$  are then used in a second-stage regression of time to healing that conditions on the same set of covariates,  $\mathbf{B}$  (the design matrix): stage 1,

$$\begin{aligned} x_i &\sim \text{Bern}(p_i^*), \\ \text{logit}(p_i^*) &= \alpha_0 + \boldsymbol{\alpha}^{\text{IV}} \cdot \mathbf{z}_i + \boldsymbol{\alpha}^{\text{adj}} \cdot \mathbf{B}_i; \end{aligned} \quad (6a)$$

stage 2,

$$y_i \sim N(y_i^*, \sigma),$$

$$y_i^* = \beta_0 + \beta^{IV} \cdot p_i^* + \beta^{adj} \cdot \mathbf{B}_i \quad \text{for } i = 1, \dots, N. \quad (6b)$$

We also explored treatment effect interactions (equations (7)). An interaction term will be endogenous (i.e. subject to confounding) in describing health outcome, even if the covariate is not itself endogenous. This means that we need to define two first-stage regressions (stage 1A and stage 1B) and, therefore, two valid instruments, one for each endogenous term, the treatment variable  $x$  and the treatment interaction variable  $x \cdot w$ . However, if  $z$  is a valid (single) instrument for the treatment variable and if the interaction variable  $w$  is exogenous then  $z \cdot w$  is a valid instrument for the interaction term (Mullahy, 1994, 1997; Bun and Harrison, 2014). The set of associated regressions when evaluating the treatment effect are then as follows (where  $\mathbf{C}$  is the design matrix): stage 1A,

$$x_i \sim \text{Bern}(p_i^a),$$

$$\text{logit}(p_i^a) = \alpha_0 + \alpha^{IV} \cdot z_i + \alpha^{IV \text{ int}} \cdot z_i \cdot w_i + \alpha^{\text{int}} \cdot w_i + \alpha^{adj} \cdot \mathbf{C}_i; \quad (7a)$$

stage 1B,

$$x_i w_i \sim \text{Bern}(p_i^b),$$

$$\text{logit}(p_i^b) = \gamma_1 + \gamma^{IV} \cdot z_i + \gamma^{IV \text{ int}} \cdot z_i \cdot w_i + \gamma^{\text{int}} \cdot w_i + \gamma^{adj} \cdot \mathbf{C}_i; \quad (7b)$$

stage 2,

$$y_i \sim N(y_i^*, \sigma),$$

$$y_i^* = \beta_1 + \beta^{IV,a} \cdot p_i^a + \beta^{IV,b} \cdot p_i^b + \beta^{\text{int}} \cdot w_i + \beta^{adj} \cdot \mathbf{C}_i, \quad \text{for } i = 1, \dots, N. \quad (7c)$$

Standard non-informative normal priors were used (not shown).

#### 2.4.2. Selection of variables for adjustment

By definition, relevant adjustment covariates are those that are associated with outcomes and are imbalanced across treatment groups, i.e. the intersection of covariates that are associated with outcomes and with treatment allocation. Given the absence of good epidemiological information in this area, and the concern that relevant predictors of healing cannot be identified on our sample because of the potential imbalances in treatment assignment, instead of using the intersection of the covariates that are associated with outcomes and treatment allocation we used their union. Although precision may be lost by the inclusion of potentially irrelevant covariates, accuracy may be gained by not missing important confounding factors.

A logistic regression with treatment received as the dependent variable was used to identify factors that are associated with treatment assignment. A thorough model selection process was devised. We started by regressing all available baseline covariates individually and selected those leading to differences of 5 units in the DIC (Lunn, 2013) in relation to the null model. We then evaluated models by using combinations of these covariates (2 by 2, 3 by 3, etc.) and selected the covariate set in the model with the best DIC.

To identify determinants of healing and potential treatment effect modifiers, we looked at the relationship of available covariates with healing by using the covariate selection process that was described above. To avoid potential confounding, we did this on participants who used NPWT and, separately, on those who did not use NPWT. To accommodate potential treatment effect modifiers we fitted treatment-by-covariate models to the full data set by using all adjustment covariates identified.



#### 2.4.3. Identification of instrumental variables $z$

A critical step in IV regression is the identification of instrument(s)  $z$ . On the basis of previous applied examples in healthcare (Brookhart *et al.*, 2006a,b), the project team defined *a priori* potential instruments which related to clinicians' practice patterns and stated preferences on the use of NPWT. The expectation was that a patient is more likely to be treated with NPWT if the clinician has a preference for it, and it was assumed that preference for NPWT is not associated with time to healing except through the choice of treatment. Specifically, we considered as a potential instrument a covariate describing whether the health professional had prescribed NPWT to their previous patient. We also surveyed the health professionals on the cohort study regarding their stated preferences for NPWT and relevant proxies (see the on-line appendix 3). Health professionals who thought that NPWT was better at managing these wounds, were less expensive or represented better value for money were assumed to be more likely to use NPWT.

Formal tests were conducted to assess the performance of each instrument (appendix 4 in the on-line supplementary material), namely for evidence of endogeneity (the Durbin–Wu–Hausman test (Greene, 2012)), evidence of the instrument set not being valid (the Hansen–Sargan test (Sargan, 1958; Hansen, 1982)), evidence of a weak instrument set (the Stock and Yogo procedure (Andrews *et al.*, 2005)) and finally evidence of model misspecification (the Pesaran–Taylor reset test (Pesaran and Taylor, 1999)).

#### 2.4.4. Extended modelling approach to consider more explicitly the 'hard-to-heal' subpopulation

In this section we broaden the analyses to ignore the imputation and instead model the cohort data in two parts: the first part describes the probability of healing within follow-up,  $p^h$ , and the second the expected time to healing for those who healed,  $x^*$ . The first part used logistic regression and the second part linear regression. Determinants of the probability of healing could differ from determinants of time to healing. We implemented adjustment on observables first (analogously to that described earlier) and then added the adjustment on unobservables by using IV regression. The implementation of the IV regression is more complex as a treatment effect is included in each equation and thus IVs are required for both.

As shown in equations (8), the predicted values for the probability of treatment allocation,  $p^h$ , are used in the two stage 2 regressions (where  $\mathbf{D}$  is the design matrix): stage 1,

$$\begin{aligned} x_i &\sim \text{Bern}(p_i^*), \\ \text{logit}(p_i^*) &= \alpha_0 + \alpha_1 \cdot z_i + \boldsymbol{\alpha}^{\text{adj}} \cdot \mathbf{D}_i; \end{aligned} \quad (8a)$$

stage 2A,

$$\begin{aligned} h_i &\sim \text{Bern}(p_i^h), \\ \text{logit}(p_i^h) &= \gamma_0 + \gamma^{\text{IV}} \cdot p_i^* + \boldsymbol{\gamma}^{\text{adj}} \cdot \mathbf{D}_i; \end{aligned} \quad (8b)$$

stage 2B,

$$\begin{aligned} y_i^h &\sim N(y_i^*, \sigma), \\ y_i^* &= \beta_0 + \beta^{\text{IV}} \cdot p_i^* + \boldsymbol{\beta}^{\text{adj}} \cdot \mathbf{D}_i, \quad \text{for } i = 1, \dots, N. \end{aligned} \quad (8c)$$

The mean time to healing was evaluated from the stage 2 equations and from expert opinion on the expected time to healing for censored observations. This was elicited according to best practice (Kahneman *et al.*, 1982; O'Hagan, 2006; Soares *et al.*, 2018; Sullivan and Payne, 2011) by asking three health professionals (who were involved in the cohort study) about their beliefs

on the additional time to 50% and 90% of patients healed (the wording is reproduced in appendix 5 in the on-line supplementary material). The elicited quantities from each expert were used to derive the parameters of a gamma distribution ( $a$  and  $b$ ) and the logarithm of the gamma parameter values were then pooled by using a standard multivariate normal formulation.

## 2.5. Inferences on health-related quality-of-life weights and costs

Given that the causal cost-effectiveness model assumes that NPWT affects overall HRQOL and costs only via effects on healing, inferences here aim to quantify the difference in HRQOL and costs between the unhealed and healed, i.e. the difference in the group means. Hence, models that are implemented here are not adjusted for further covariates, but these are implicitly averaged over so that the results can be considered to be representative of the population that is included in the cohort study.

Observations of EuroQol 5 dimensions, EQ-5D, that were collected within the cohort study at baseline and every 3 months thereafter up to 18 months follow-up were used. The EQ-5D index score was computed by using a published valuation algorithm (Brooks and De Charro, 1996). A panel data approach (a mixed model with random intercept) to capture both within- and across-patient variation was used to model change in the index score from baseline, using a time-dependent indicator of whether the patient had healed or not as the only covariate. Exchangeability across time within individuals was assumed. Flat uniform priors were used for variance parameters, and others used vague priors (which are not shown here). The estimated effect of healing on HRQOL was used to inform directly the term  $u_h - u_{\text{unh}}$  in the causal cost-effectiveness model in equation (1).

As with EQ-5D, resource use was collected within the cohort study at baseline and at 3-monthly intervals, and included general practitioner visits, nurse visits, outpatient visits and hospital admissions. Costs per patient were obtained by multiplying resource use by relevant unit costs (2014 prices). Unit costs were obtained from the Personal Social Services Research Unit, 2014 (Curtis, 2014), and National Health Service reference costs 2013–2014 (Department of Health, 2014). Treatment costs were obtained from the British National Formulary (2014) as well as specific costing information from the two centres (Hull and Leeds) on NPWT. A detailed listing of the specific values of unit costs is shown in appendix 2, Table A2.1, in the on-line supplementary material.

Many observations over the different time points were expected to have zero costs, and hence a Bayesian two-part mixed model was used: the first part captured the probability of zero costs and the second part the costs of those with non-zero values. A mixed gamma model with log-link was used because of data skewness. Note that costs for patients healed are not necessarily expected to be zero as time to healing collected within the cohort study refers to an index wound and the patient may have other wounds.

## 3. Results

### 3.1. Effectiveness

#### 3.1.1. Using the complete data set

Wounds were more likely to be treated with NPWT if the wound was larger, if there was more tissue involvement and if the patient was an inpatient (the results are not shown here). Within at least one treatment group, wound history ('yes' versus 'no' to having had an SWHSI before) and treatment location (inpatient versus outpatient) were associated with healing. Of these, wound history was identified as a treatment effect modifier.

**Table 1.** Effectiveness results: complete data set†

<i>Results for the following models:</i>					
	<i>Model 0</i>	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>	<i>Model 4</i>
Instrument used, <i>z</i>	None	None	None	Previous treatment	Previous treatment with interaction
Equation	Time to healing	Time to healing	Time to healing	(Stage 2) time to healing	(Stage 2) time to healing
Adjustment covariates	None	Observables	Observables + interaction	Observables	Observables + interaction
Area (> 25 cm <sup>2</sup> )‡	—	18.0 [−23.1, 58.0]	21.1 [−20.3, 62.1]	23.8 [29.8, 43.7]	20.7 [−42.9, 80.4]
Treatment location§	—	45.3 [11.3, 79.2]	43.8 [9.9, 76.8]	48.9 [3.0, 63.8]	48.2 [−2.7, 100.5]
Tissue involvement§§	—	31.9 [−1.5, 65.0]	26.6 [−5.3, 58.7]	33.8 [−8.0, 73.1]	31.0 [−11.1, 71.7]
History	—	7.5 [−28.4, 42.9]	−29.3 [−71.5, 12.1]	7.2 [−29.5, 44.6]	−9.1 [−62.3, 39.3]
NPWT	108.0 [73.6, 142.2]	73.2 [33.8, 112.8]	41.9 [−1.1, 84.1]	56.9 [−71.7, 192.8]	53.8 [−81.9, 199.8]
NPWT × History	—	—	138.7 [56.9, 221.8]	—	52.4 [−70.4, 212]
Constant	116.8 [97.9, 135.2]	81.4 [53.3, 110.0]	92.7 [64.5, 120.8]	81.1 [52.3, 109.1]	88.1 [57.5, 118.6]
Observations	393	372	372	372	372

†Imputation using minimum date of healing.

‡Wound area above 25 cm<sup>2</sup> (*versus* wound area below 25 cm<sup>2</sup>).

§Treatment location: inpatient (*versus* outpatient).

§§Tissue involvement, skin and subcutaneous tissue loss (*versus* skin loss).

Effectiveness results under the complete, imputed, time to healing data are summarized in Table 1: model 0 is unadjusted (presented for completeness), model 1 adjusts for observables but excludes the interaction term and model 2 includes wound history as an interaction term. The results show that, after adjustment, patients using NPWT are still expected to take longer to heal than those who do not use NPWT. Model 1 estimates an increase in time to healing of 73.2 days: a statistically significant result. Model 2 estimates that patients with a history of SWHSI take 181 days longer to heal with NPWT, whereas patients without a history of SWHSI took 42 days longer to heal with NPWT.

Regarding the IV approach, previous treatment used by the treating health professional was the variable that performed best (see appendix 4 in the on-line supplementary material). Using this instrument, however, still led to higher time to healing estimates (but not statistically significant) for patients using NPWT—Table 1, models 3 and 4.

### 3.1.2. Extended analysis

The determinants of the probability of healing (stage 2A, equation (8)) identified were the body mass index BMI and wound location. For time to wound healing, determinants identified were surgery type (colorectal *versus* other surgeries), wound duration and diabetic foot wounds (stage 2B, equation (8c)). No evidence of relevant interaction terms was found.

**Table 2.** Time to healing: extended modelling approach†

<i>Results for the following models:</i>				
	<i>Model b0</i>		<i>Model b1</i>	
Equation	(Stage 2A) <i>P</i> (heal) (odds ratio)	(Stage 2B) time to healing	(Stage 2A) <i>P</i> (heal) (odds ratio)	(Stage 2B) time to healing
Type of regression model	Logit	Ordinary least squares	IV, logit	IV, ordinary least squares
Adjustment	None	None	Adjustment set	Adjustment set
Area (> 25 cm <sup>2</sup> )‡	—	—	0.52	27.7
	—	—	[0.24, 0.99]	[0.6, 54.5]
Treatment location§	—	—	0.52	30.9
	—	—	[0.22, 1.04]	[9.2, 52.3]
Tissue involvement§	—	—	0.50	11.4
	—	—	[0.25, 0.91]	[−8.1, 30.9]
BMI	—	—	1.04	—
	—	—	[1.00, 1.09]	—
Location—foot	—	—	0.34	30.1
	—	—	[0.16, 0.64]	[9.2, 52.3]
Surgery type	—	—	—	13.1
	—	—	—	[−6.1, 32.4]
SWHSI duration	—	—	—	2.3
	—	—	—	[0.6, 4.0]
Diabetic feet	—	—	—	42.8
	—	—	—	[9.1, 76.3]
NPWT	0.35	69.0	0.59	46.0
	[0.20, 0.57]	[48.2, 89.1]	[0.28, 1.12]	[19.6, 72.5]
Constant	6.56	82.0	6.98	36.9
	[4.70, 9.37]	[71.9, 92.1]	[1.62, 29.55]	[12.4, 61.3]
Observations	393		354	

† $x_1, x_2, x_3, v_1, v_2, v_4$  and  $v_5$  are adjustment covariates;  $w$  is the covariate interacted with treatment;  $t$  is the treatment covariate;  $w \cdot t$  is the treatment interaction;  $z$  is the instrument.

‡Wound area above 25 cm<sup>2</sup> (*versus* wound area below 25 cm<sup>2</sup>).

§Treatment location: inpatient (*versus* outpatient).

The extended model predicted that NPWT was associated with poorer healing outcomes (46 expected additional days to wound healing compared with the non-NPWT patients), using previous treatment used by the health professional as the instrument (Table 2, model b1). This result was not statistically significant.

### 3.2. Modelling of health-related quality-of-life weights and costs

There were 1484 observations of the EQ-5D index score over all time points and 1166 observations of disease-related costs (614, or 53.7%, had a zero value of costs). A descriptive summary of results is shown in appendix 2, Table A2.2, of the on-line supplementary material. The formal modelling of the EQ-5D index score (Table 3) shows that the effect of healing is small (an increase of 0.055 in the index score for those who healed compared with those who did not) but statistically significant (the credible interval does not include zero).

The modelling of costs shows that, overall, expected costs for unhealed and healed patients were respectively £1124.4 and £258.9 per quarter (Table 3)—a difference of £865 attributed to

**Table 3.** EQ-5D index score and disease-related costs (regression analysis)

<i>Model</i>	<i>Mean (£)</i>	<i>95% credible interval</i>	
Expected EQ-5D, unhealed	0.091	0.036	0.146
Expected EQ-5D, healed	0.091 + 0.055 = 0.146	0.098	0.193
Difference in EQ-5D attributed to healing	0.055	0.015	0.094
Expected costs, unhealed	1124.45	916.9	1359.1
Expected costs, healed	258.91	208.1	317.7
Difference in costs attributed to healing	865	674.2	1126.5

**Table 4.** Elicited evidence on additional time to healing from follow-up for censored patients

	<i>Elicited evidence</i>		<i>Fitted values for parameters of a gamma distribution</i>				
	<i>50th percentile (days)</i>	<i>90th percentile (days)</i>	<i>Shape, 1st and 3rd quartiles</i>	<i>Scale, 1st and 3rd quartiles</i>	<i>Mean (days), 1st and 3rd quartiles</i>	<i>50th percentile, (days), 1st and 3rd quartiles</i>	<i>90th percentile, (days), 1st and 3rd quartiles</i>
Expert 1	1044	3234	1.12	1290	1445	1031	3322
Expert 2	496	679	15.3	33.1	506	495	683
Expert 3	180	3650	0.23	5222	1201	176	3176
Pooled using formal synthesis [95% credible interval]	—	—	1.79 [0.03, 54.5]	615 [241, 1615]	1038 [29, 39065]	559 [0, 28011]	1820 [15, 37161]

healing (the standard error over the difference was £111.9). Further details can be found in appendix 2, Tables A2.3 and A2.4, in the on-line supplementary material.

The estimated patient time on NPWT at any stage was 36.9 days on average (standard error SE = 6.0). The average per-patient cost of NPWT treatment is thus evaluated at £1180.0 (SE = £189.2).

The responses to the elicitation exercise highlight that experts believe that up to 10% of these hard-to-heal patients may be unhealed 2–10 years after having exited the cohort study (Table 4). The pooled gamma distribution across the three experts determines a median additional time to healing of 1038.5 days. In the cost-effectiveness model, time to healing was replaced by the patient’s life expectancy when patients were predicted to heal after their life expectancy. Considering the gender split in our sample (56.4% were male), the life expectancy for the sample is 83.9 years of age (Office for National Statistics, 2014).

Cost-effectiveness results, for the complete data set analyses, indicate that NPWT is less effective and more costly, and thus would not be recommended for use in SWHSI patients—Table 5. The results show no decision uncertainty, indicating that the confidence in the decision, based on the data and assumptions of analyses, is high and that further research is unlikely to change this recommendation.

**Table 5.** Cost-effectiveness results†

	Results for complete data set						Extended analysis, model b1
	Model 1	Model 2		Model 3	Model 4		
		With history	Without history		With history	Without history	
Incremental QALYs	−0.012 [0.005]	−0.029 [0.012]	−0.007 [0.004]	−0.008 [0.011]	−0.013 [0.017]	−0.001 [0.011]	−0.027 [0.017]
Incremental costs	1979 [316]	3573 [534]	1592 [330]	1749.4 [809.8]	2144.8 [1210.8]	1187.0 [844.9]	3116.1 [1025.9]
ICER	Dominated	Dominated	Dominated	Dominated	Dominated	Dominated	Dominated
Probability CE(£20000)†	0	0	0	0.026	0.053	0.123	0.002
Probability CE(£30000)†	0	0	0	0.032	0.060	0.144	0.003

†Probability that NPWT is cost effective at £20000/QALY or £30000/QALY cost-effectiveness thresholds. Standard errors are in brackets.

The extended analyses indicated that for average patient characteristics (BMI = 29; diabetes = 26%; cardiovascular or peripheral disease = 43%; smoker = 29%; wound area bigger than 25 cm<sup>2</sup> = 24%; emergency surgery = 64%; location other than abdomen = 66%; open wound planned = 41%; history of SWHSI = 25%) those treated with NPWT would be expected to heal in 251 days whereas those not treated with NPWT would be expected to heal in 137 days. The results indicate that NPWT is not cost effective for the treatment of SWHSI and, again, there is very little uncertainty over this result (Table 5, model b2).

#### 4. Discussion

Recent policy developments, in the UK and elsewhere, are increasingly establishing the use of observational data to support decisions about the use of healthcare technologies, particularly where significant clinical uncertainties remain. This paper uses an applied example where data from an observational cohort study are used to evaluate the clinical and cost-effectiveness of NPWT in the healing of SWHSI. We explored different ways of adjusting for confounding and implemented adjustment for observables (using regression) and unobservables (using IV regression). But the different modelling approaches produced similar results, all suggesting that NPWT is neither effective nor cost effective compared with standard dressings. The IV approach reduced the mean additional days to heal for NPWT patients compared with other treatments from 112 (95% credible interval [94,141]) (unadjusted estimate) to 57 days with a wide credible interval (95% credible interval [−72, 193]), potentially better reflecting the existing uncertainty surrounding the effectiveness of NPWT.

Our analysis identified many challenges to the analysis of observational data that have not been hitherto considered in guidance. A first relates to core definitions, such as of outcomes that are affected by treatment or the definition of treatments themselves. Our analyses focused on time to wound healing: the outcome that is considered clinically most relevant (Cullum *et al.*, 2015; Dumville *et al.*, 2015). However, treatment may also affect other outcomes—it may reduce the rate of wound infection or facilitate wound management—and failing to identify the relevant

outcomes of treatment can misinform recommendations. In what concerns the definition of the treatments, given the limited sample size our study pooled patients who received NPWT at baseline and patients who received NPWT at a later stage (89 and 26 respectively). However, patients receiving NPWT at inception can be argued to differ (and to have a different prognosis) from those who received NPWT during follow-up. There may also be a potential bias from the presence of time-dependent confounding, as patients who start NPWT after baseline are likely to have been switched to NPWT as a result of previous treatment failure. In addition, we have merged the different dressings that were used within the cohort into a single comparator (the 'non-NPWT' group). This intervention 'lumping' may be a source of further heterogeneity. All analyses of observational data are likely to have to make similar decisions on practical grounds. Within observational data sets, heterogeneity is expected to be more pronounced than in randomized controlled trials that define clear inclusion and exclusion criteria, determine the intervention and how it is applied and often determine follow-on care. In our case-study, heterogeneity in the patients who were recruited was extensive, e.g. in what concerns the location of the wound, the surgery that led to the wound, its size at baseline and the setting of care. However, the small sample size ( $n = 393$ ) was insufficient to explore the consequences of existing heterogeneity, i.e. to explore the implications of treatment within potentially relevant subgroups of the population (such as open abdominal wounds *versus* diabetic foot and leg wounds).

A second aspect relates to the potential for the data set to show a lack of overlap, where regions of the covariate space have relatively few treated or control units. If there is selection into treatment, a lack of overlap (to a bigger or lesser extent) is expected. Inferences that are made for such regions rely on extrapolation (Imbens and Rubin, 2015), with results being more dependent on model specification and less on direct support from the data (Gelman and Hill, 2007). This is relevant for any form of adjustment, whether just on observed characteristics and/or on unobserved characteristics through IV regression. Although some overlap is required to enable meaningful analyses (Faria *et al.*, 2015b) there is no clear guidance on how to establish when overlap is a significant problem. Further research is needed to understand how alternative methods (e.g. regression adjustment, matching or inverse probability weighting) could capitalize on the underlying overlap issue (Gelman and Hill, 2007). Our case-study highlighted that, in practice, it may even be difficult to identify the sources of lack of overlap, especially when no single covariate arises as meaningful.

A third aspect relates to the adjustment on observables. The fact that health outcomes and their determinants were not known *a priori* in our case-study, and that the mechanism for selection was also not known, meant that it was difficult to identify the appropriate adjustment set. In terms of covariate selection, we opted for an information theoretic approach and considered models with different combinations of covariates to identify the best fitting one according to the DIC. Because these methods are computationally intensive (even for a relatively small number of covariates), within this study we developed a thorough and protocol-driven variable-selection process that can be generalized and used in any other application. There are various alternative approaches to variable selection in the literature, and further research could explore these in the context of this case-study. We highlight the Bayesian spike and slab, that defines mixture priors for the covariate effects (Bayarri *et al.*, 2012; O'Hara and Sillanpää, 2009) and hence integrates uncertainty over covariate selection in the model estimates. In terms of methods of adjustment for observables, the use of an adjustment regression approach facilitated the further application of the IV method to address unobserved confounding. Although there is a range of alternative methods of adjustment, such as matching approaches, their relative performance is as yet unclear (Faria *et al.*, 2015; Kreif *et al.*, 2013), particularly with small sample sizes and poor overlap.

Another set of challenges relates to the application of IV techniques to control for unobserved confounding. Such methods depend on finding good and reliable, but difficult to validate, sources of exogenous variation (instruments) (Faria *et al.*, 2015). More research is needed in defining instruments that are relevant in the context of health technology assessment, and on extending analyses to outcomes such as time-to-event data.

Also, in our case-study we used a Bayesian context for inferences, which we believe has clear advantages. It enables flexibility in specifying the model according to what is believed to be plausible without being limited to linearity or normality assumptions. Although in this work we used linearity and normality to describe time to healing, an important extension for future work is either to consider a log-normal transformation or to use time-to-event distributions and to explore potential non-linearity. Another important advantage is where the sample is small or the instruments weakly related to treatment assignment, in which case Bayesian methods are thought to return more accurate estimates of causal effects (Crespo-Tenorio and Montgomery, 2013). This is crucial, as in this area of research it is likely that the sample size is limited. However, Bayesian research on IV models is limited and recent, relative to the broad literature from the classical (i.e. frequentist) statistical perspective. An informal search (without date restriction) identified only 24 methodological papers focusing on aspects of Bayesian IV regression—one provides an overview (Kleibergen and Zivot, 2003), 16 develop Bayesian estimators for particular model specifications (Hollenbach *et al.*, 2019; Kato and Hoshino, 2018; Shi and Tong, 2017; Li and Lu, 2015; Lopes and Polson, 2014; Crespo-Tenorio and Montgomery, 2013; Kato, 2013; Zellner *et al.*, 2014; Chamberlain and Imbens, 1996, 2003; Wiesenfarth *et al.*, 2014; Kobayashi and Ogasawara, 2016; National Cancer Registration and Analysis Service, 2016; Conley *et al.*, 2008, 2012; Burgess *et al.*, 2010), one focuses on variable selection (Sabnis *et al.*, 2019) and six consider aspects of the validity of instruments (Kraay, 2012; Koop *et al.*, 2012; Eicher *et al.*, 2009; Lenkoski *et al.*, 2014; Karl and Lenkoski, 2018; Henry *et al.*, 2018). It is important to highlight that none provided generic software code of the model or for testing the validity of instruments.

Future methodological research is needed to help to guide the many practical challenges that are faced. This is important for analysts, but crucial for policy as it may improve the confidence that decision makers place in the results of such analyses. Even with our case-study indicating that NPWT should not be recommended for the healing of SWHSI and that further research is unlikely to change this recommendation, the National Institute for Health Research in the UK funded a follow-on pilot trial (which is likely to report in 2022). Such a lack of confidence in analyses of observational data may become a problem when these types of analyses are to be used more widely to inform system level decisions such as those required within the Cancer Drugs Fund (Grieve *et al.*, 2016).

## 5. Declarations

### 5.1. Ethics approval and consent to participate

The cohort study received research ethics (Research Ethics Committee reference: 11/YH/0313) and written informed consent was obtained from all participants before participation.

Before beginning the cohort study, approval was obtained from Yorkshire and Humber–Humber Bridge Research Ethics Committee (reference: 12/YH/0350) and from the associated National Health Service Trusts.

Before study activity began, approval was obtained from Yorkshire and Humber–Leeds East Research Ethics Committee (reference: 15-YH-0307) and from the associated National Health Service Trusts.



## 5.2. Availability of data and materials

The data sets that were used and/or analysed during the current study are available from the corresponding author on reasonable request.

## 5.3. Competing interests

Professor Claxton reports personal fees from F. Hoffmann-La Roche Ltd, personal fees from Heron (PAREXEL), outside the submitted work. Dr Welton reports grants from the National Institute for Health Research during the conduct of the study; in particular, she was supported by the National Institute for Health Research Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol; she is also Principal Investigator for a research grant jointly funded by the Medical Research Council and Pfizer.

The views expressed in this publication are those of the author(s) and not necessarily those of the National Health Service, the National Institute for Health Research or the Department of Health and Social Care.

All the other authors have no conflicts of interest to declare.

## 5.4. Authors' contributions

Dr Pedro Saramago (Research Fellow, Health Economics) undertook all the analyses and contributed to the design, conduct, interpretation and reporting. Professor Karl Claxton (Professor, Economics) oversaw all aspects of work, including design, conduct, analysis, interpretation and reporting. Professor Nicky Welton (Professor, Statistics and Health Economic Modelling) oversaw the design, interpretation and reporting of the analyses. Dr Marta O. Soares (Senior Research Fellow, Health Economics) supervised all aspects of the work, including design, conduct, analysis, interpretation and reporting.

## Acknowledgements

We acknowledge Richard Grieve, Professor of Health Economics Methodology from the London School of Hygiene and Tropical Medicine, and participants in the 2016 Health Economists' Study Group winter conference in Manchester (UK) for their very helpful comments on an earlier draft of this work. We also acknowledge Professor Jo Dumville, Professor Mike Drummond and Professor Mark Sculpher for their comments and guidance along the way. Finally, we acknowledge all SWHSI research team members for their useful comments and suggestions.

Funding came from a grant from the National Institute for Health Research programme grants for applied research (project RP-PG-0609-10171). The funding agreement ensured the authors' independence in designing the study, interpreting the data and writing and publishing this paper.

## References

- Akhmetov, I. and Bubnov, R. V. (2015) Assessing value of innovative molecular diagnostic tests in the concept of predictive, preventive, and personalized medicine. *EPMA J.*, **6**, article 19.
- Andrews, D. W. K., Stock, J. H. and Rothenberg, T. J. (2005) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press.
- Banwell, P. E. and Musgrave, M. (2004) Topical negative pressure therapy: mechanisms and indications. *Int. Wound J.*, **1**, 95–106.
- Bayarri, M. J., Berger, J. O., Forte, A. and Garcia-Donato, G. (2012) Criteria for Bayesian model choice with application to variable selection. *Ann. Statist.*, **40**, 1550–1577.
- Berger, M. L., Martin, B. C., Huserau, D., Worley, K., Allen, J. D., Yang, W., Quon, N. C., Mullins, C. D.,

- Kahler, K. H. and Crown, W. (2014) A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Val. Hlth*, **17**, 143–156.
- British National Formulary (2014) British National Formulary–Joint Formulary Committee. British Medical Association and Royal Pharmaceutical Society of Great Britain, London.
- Brookhart, M. A., Wang, P. S., Solomon, D. H. and Schneeweiss, S. (2006a) Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology*, **17**, 268–275.
- Brookhart, M. A., Wang, P. S., Solomon, D. H. and Schneeweiss, S. (2006b) Instrumental variable analysis of secondary pharmacoepidemiologic data. *Epidemiology*, **17**, 373–374.
- Brooks, R. and De Charro, F. (1996) EuroQol: the current state of play. *Hlth Poly*, **37**, 53–72.
- Bun, M. J. G. and Harrison, T. D. (2014) OLS and IV estimation of regression models including endogenous interaction terms. *Econometr. Rev.*, **38**, no. 7.
- Burgess, S., Thompson, S. G. and CRP CHD Genetics Collaboration (2010) Bayesian methods for meta-analysis of causal relationships estimated using genetic instrumental variables. *Statist. Med.*, **29**, 1298–1311.
- Byron, S. K., Crabb, N., George, E., Marlow, M. and Newland, A. (2014) The Health Technology Assessment of companion diagnostics: experience of NICE. *Clin. Cancer Res.*, **20**, 1469–1476.
- Cameron, A. C. and Trivedi, P. K. (2005) *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.
- Chamberlain, G. and Imbens, G. W. (1996) Hierarchical Bayes models with many instrumental variables. *Working Paper*. National Bureau of Economic Research, Cambridge.
- Chamberlain, G. and Imbens, G. W. (2003) Nonparametric applications of Bayesian inference. *J. Bus. Econ. Statist.*, **21**, 12–18.
- Chetter, I. C., Oswald, A. V., McGinnis, E., Stubbs, N., Arundel, C., Buckley, H., Bell, K., Dumville, J. C., Cullum, N. A., Soares, M. O. and Samago, P. (2019) Patients2017: a prospective, observational cohort study of patients with surgical wounds healing by secondary intention: a prospective, cohort study. *Int. J. Nurs. Stud.*, **89**, 62–71.
- Conley, T. G., Hansen, C. B., McCulloch, R. E. and Rossi, P. E. (2008) A semi-parametric Bayesian approach to the instrumental variable problem. *J. Econometr.*, **144**, 276–305.
- Conley, T. G., Hansen, C. B. and Rossi, P. E. (2012) A non-parametric Bayesian approach to the instrumental variable problem. In *Review of Economics and Statistics*. Cambridge: MIT Press.
- Crespo-Tenorio, A. and Montgomery, J. (2013) A Bayesian approach to inference with instrumental variables: improving estimation of treatment effects with weak instruments and small samples. Department of Political Science, Washington University, St Louis. (Available from <https://pages.shanti.virginia.edu/PolMeth/files/2013/07/CrespoTenorioMontgomery.pdf>.)
- Cullum, N., Buckley, H. L., Dumville, J., Hall, J., Lamb, K., Madden, M. T., Morley, R., O'Meara, S. M., Samago Goncalves, P. R., Soares, M. O. and Stubbs, N. (2015) Wounds research for patient benefit: a 5 year programme of research. *Programme Grant for Applied Research 4.13*. National Institute for Health Research Journals Library, Southampton.
- Curtis, L. (2014) Unit costs of health & social care 2014. Personal Social Services Research Unit, University of Kent, Canterbury.
- Department of Health (2011) Department of Health: co-dependent and hybrid technologies. Department of Health, Australian Government, Canberra.
- Department of Health (2014) Reference costs 2013–2014. NHS England, London.
- Dumville, J. C., Owens, G., Webster, J., Evans, D., Land, L., Crosbie, E. J., Peinemann, F. and Liu, Z. (2015) Negative pressure wound therapy for treating surgical wounds healing by secondary intention: pressure ulcers. In *Cochrane Database Systematic Reviews*, vol. 6, CD0112785, CD011334.
- Eicher, T. S., Lenkoski, A. and Raftery, A. E. (2009) Bayesian model averaging and endogeneity under model uncertainty: an application to development determinants. *Working Paper*. Center for Statistics and the Social Sciences, University of Washington, Seattle.
- European Medicines Agency (2020) Medical devices. European Medicines Agency, London. (Available from [https://www.ema.europa.eu/en/human-regulatory/overview/medical-devices#companion-diagnostics-\(in-vitro-diagnostics\)-section](https://www.ema.europa.eu/en/human-regulatory/overview/medical-devices#companion-diagnostics-(in-vitro-diagnostics)-section).)
- Faria, R., Hernandez Alava, M., Manca, A. and Wailoo, A. (2015) The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: methods for comparative individual patient data. *Technical Support Document 17*. National Institute for Health and Care Excellence, London.
- Gelman, A. (2014) *Bayesian Data Analysis*. Boca Raton: CRC Press.
- Gelman, A. and Hill, J. (2007) *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Greene, W. H. (2012) *Econometric Analysis*. Boston: Prentice Hall.
- Grieve, R., Abrams, K., Claxton, K., Goldacre, B., James, N., Nicholl, J., Parmar, M., Parker, C., Sekhon, J. S., Smeeth, L., Spiegelhalter, D. and Sculpher, M. (2016) Cancer Drugs Fund requires further reform. *Br. Med. J.*, **354**, article i5090.
- Hadley, J., Polsky, D., Mandelblatt, J. S., Mitchell, J. M., Weeks, J. C., Wang, Q., Hwang, Y. T. and OPTIONS

- Research Team (2003) An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population. *Hlth Econ.*, **12**, 171–186.
- Hansen, L. P. (1982) Large sample properties of generalized-method of moments estimators. *Econometrica*, **50**, 1029–1054.
- Henry, T. R., Fisher, Z. F. and Bollen, K. A. (2020) Bayesian model averaging for model implied instrumental variable two stage least squares estimators. *Preprint arXiv:1808.10522*. Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill.
- Henshall, C., Mardhani-Bayne, L., Fronsdaal, K. B. and Klemp, M. (2011) Interactions between health technology assessment, coverage, and regulatory processes: emerging issues, goals, and opportunities. *Int. J. Technol. Assessmnt Hlth Care*, **27**, 253–260.
- Hollenbach, F. M., Montgomery, J. M. and Crespo-Tenorio, A. (2019) Bayesian versus maximum likelihood estimation of treatment effects in bivariate probit instrumental variable models. *Polit. Sci. Res. Meth.*, **7**, 651–659.
- Imbens, G. and Rubin, D. B. (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences: an Introduction*. New York: Cambridge University Press.
- Kahneman, D., Slovic, P. and Tversky, A. (1982) *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Karl, A. and Lenkoski, A. (2018) Instrumental variable Bayesian model averaging via conditional Bayes factors. *Preprint arXiv:1202.5846*. Heidelberg University, Heidelberg.
- Kato, K. (2013) Quasi-Bayesian analysis of nonparametric instrumental variables models. *Ann. Statist.*, **41**, 2359–2390.
- Kato, R. and Hoshino, T. (2018) Semiparametric Bayes instrumental variable estimation with many weak instruments. *Discussion Paper Series*. Research Institute for Economics and Business Administration, Kobe University, Kobe.
- Kleibergen, F. and Zivot, E. (2003) Bayesian and classical approaches to instrumental variable regression. *J. Econometr.*, **114**, 29–72.
- Kobayashi, G. and Ogasawara, K. (2016) Bayesian nonparametric instrumental variable regression approach to quantile inference. *Preprint arXiv:1608.07921*. Faculty of Law, Politics and Economics, Chiba University, Chiba.
- Koop, G., Leon-Gonzalez, R. and Strachan, R. (2012) Bayesian model averaging in the instrumental variable regression model. *J. Econometr.*, **171**, 237–250.
- Kraay, A. (2012) Instrumental variables regressions with uncertain exclusion restrictions: a Bayesian approach. *J. Appl. Econometr.*, **27**, 108–128.
- Kreif, N., Grieve, R. and Sadique, M. Z. (2013) Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. *Hlth Econ.*, **22**, 486–500.
- Leaper, D., Burman-Roy, S., Palanca, A., Cullen, K., Worster, D., Gautam-Aitken, E., Whittle, M. and Guide Development Group (2008) Prevention and treatment of surgical site infection: summary of NICE guidance. *Br. Med. J.*, **337**, article a1924.
- Lenkoski, A., Eicher, T. S. and Raftery, A. E. (2014) Two-stage Bayesian model averaging in endogenous variable models. *Econometr. Rev.*, **33**, 122–151.
- Li, G. and Lu, X. Y. (2015) A Bayesian approach for instrumental variable analysis with censored time-to-event outcome. *Statist. Med.*, **34**, 664–684.
- Lopes, H. F. and Polson, N. G. (2014) Bayesian instrumental variables: priors and likelihoods. *Econometr. Rev.*, **33**, 100–121.
- Lunn, D. (2013) *The BUGS Book: a Practical Introduction to Bayesian Analysis*. Boca Raton: CRC Press.
- Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statist. Comput.*, **10**, 325–337.
- Motheral, B., Brooks, J., Clark, M. A., Crown, W. H., Davey, P., Hutchins, D., Martin, B. C. and Stang, P. (2003) A checklist for retrospective database studies—report of the ISPOR Task Force on Retrospective Databases. *Val. Hlth*, **6**, 90–97.
- Mullahy, J. (1994) Instrumental variable estimation of Poisson regression models. *Proc. Bus. Econ. Statist. Sect. Am. Statist. Ass.*, 7–13.
- Mullahy, J. (1997) Instrumental-variable estimation of count data models: applications to models of cigarette smoking behavior. *Rev. Econ. Statist.*, **79**, 586–593.
- National Cancer Institute (2019) Surveillance, Epidemiology, and End Results Program. National Cancer Institute, Rockville. (Available from <https://seer.cancer.gov/>.)
- National Cancer Registration and Analysis Service (2016) SACT systemic anti-cancer therapy dataset. National Cancer Registration and Analysis Service. (Available from <http://www.chemodataset.nhs.uk/home>.)
- National Institute for Health and Care Excellence (2008) Surgical site infections: prevention and treatment. National Institute for Health and Care Excellence, London.
- National Institute for Health and Care Excellence (2011a) Medical technologies guidance. National Institute for Health and Care Excellence, London. (Available from <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-medical-technologies-guidance>.)

- National Institute for Health and Care Excellence (2011b) Review of Clinical Guideline (CG74)—Prevention and treatment of surgical site infection. National Institute for Health and Care Excellence, London.
- National Institute for Health and Care Excellence (2013) Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence, London.
- National Institute for Health Research (2011) Surgical wounds healing by secondary intention: characterising and quantifying the problem and identifying effective treatments. National Institute for Health Research, London. (Available from <http://www.nihr.ac.uk/funding/funded-research/funded-research.htm?postid=2194>.)
- NHS England (2019) NHS England—Cancer Drugs Fund. NHS England, London. (Available from <https://www.england.nhs.uk/cancer/cdf/>.)
- Office for National Statistics (2014) *ONS Statistics for England and Wales in 2011-2013*. Newport: Office for National Statistics.
- O'Hagan, A. (2006) *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester: Wiley.
- O'Hara, R. B. and Sillanpää, M. J. (2009) A review of Bayesian variable selection methods: what, how and which. *Bayes Anal.*, **4**, 85–117.
- Pesaran, M. H. and Taylor, L. W. (1999) Diagnostics for IV regressions. *Oxf. Bull. Econ. Statist.*, **61**, 255–281.
- Prentice, J. C., Conlin, P. R., Gellad, W. F., Edelman, D., Lee, T. A. and Pizer, S. D. (2014) Capitalizing on prescribing pattern variation to compare medications for type 2 diabetes. *Val. Hlth*, **17**, 854–862.
- R Core Development Team (2013) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Sabnis, G., Atchadé, Y. and Dovonon, P. (2019) Bayesian variable selection in linear regression models with instrumental variables. *Preprint arXiv:1901.03182*. Boston University, Boston.
- Sargan, J. D. (1958) The estimation of economic relationships using instrumental variables. *Econometrica*, **26**, 393–415.
- Schneeweiss, S., Seeger, J. D., Landon, J. and Walker, A. M. (2008) Aprotinin during coronary-artery bypass grafting and risk of death. *New Engl. J. Med.*, **358**, 771–783.
- Shi, D. and Tong, X. (2017) Bayesian two-stage robust causal modeling with instrumental variables using Student's *t* distributions. In *Bayesian Inference* (ed. J. P. Tejedor).
- Soares, M. O., Sharples, L., Morton, A., Claxton, K. and Bojke, L. (2018) Experiences of structured elicitation for model-based cost-effectiveness analyses. *Val. Hlth*, **21**, 715–723.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B*, **64**, 583–616.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2014) The deviance information criterion: 12 years on. *J. R. Statist. Soc. B*, **76**, 485–493.
- StataCorp (2013) *Stata Statistical Software: Release 13*. College Station: StataCorp.
- Sullivan, W. and Payne, K. (2011) The appropriate elicitation of expert opinion in economic models: making expert data fit for purpose. *Pharmacoeconomics*, **29**, 455–459.
- Wiesenfarth, M., Hisgen, C. M., Kneib, T. and Cadarso-Suarez, C. (2014) Bayesian nonparametric instrumental variables regression based on penalized splines and Dirichlet process mixtures. *J. Bus. Econ. Statist.*, **32**, 468–482.
- Wooldridge, J. M. (2013) *Introductory Econometrics: a Modern Approach*. Mason: South-Western Cengage Learning.
- World Health Organization (2011) Health technology assessment of medical devices. World Health Organization, Geneva.
- Zellner, A., Ando, T., Basturk, N., Hoogerheide, L. and Van Dijk, H. K. (2014) Bayesian analysis of instrumental variable models: acceptance-rejection within direct Monte Carlo. *Econometr. Rev.*, **33**, 3–35.

#### Supporting information

Additional 'supporting information' may be found in the on-line version of this article.